Appendix A-Algorithmic Evaluation Methodology

A.1 R^2

R^2 is an indicator for assessing the goodness of fit of the model, reflecting the proportion of the total variation in the dependent variable explained by the independent variables. The closer it is to 1, the better the fit of the model.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad \text{(A1)}$$

Where, y_i represents the actual value of equilibrium temperature, $\hat{y}_i$ represents the predicted value of equilibrium temperature, and $\bar{y}$ represents the average value of actual equilibrium temperature.

A.2 RMSE

RMSE is a measure of the difference between the predicted and true values of a model. It represents the standard deviation between the predicted and true values and is used to respond to the absolute amount of model prediction error.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad \text{(A2)}$$

Where *m* is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

A.3 MAPE

MAPE is a metric that indicates the relative error between the predicted and true values, given as a percentage, reflecting the average prediction error rate of the model. It can be written as:

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad \text{(A3)}$$

Where *m* is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

Appendix B

Table S1 Descriptive statistics of the complete feature set

	Count	Mean	Median	Std	Min	Max
T (K)	3807	349.5100	344.300	112.0800	88.7100	623.150
P (bar)	3807	73.9600	60.000	57.0400	1.0000	299.100
x	3807	0.0580	0.038	0.0640	0.0001	0.501

MW (g/mol)	3807	116.2957	88.106	107.5769	16.0430	647.252
TB (K)	3807	390.4972	383.750	153.5760	111.6700	847.530
PC (bar)	3807	37.9015	39.200	17.0687	4.5400	82.000
pc (g/mL)	3807	0.2599	0.256	0.0481	0.1627	0.381
ZC	3807	0.2501	0.259	0.0308	0.1340	0.294
w	3807	0.4274	0.379	0.3097	0.0120	1.597
numHacc	3807	0.4605	0.000	0.8443	0.0000	5.000
sa	3807	17.8585	12.000	15.3177	1.2000	83.750
psa_div_sa	3807	0.2388	0.000	0.3661	0.0000	0.906
Aldehyde	3807	0.0096	0.000	0.0974	0.0000	1.000
Ketone	3807	0.0508	0.000	0.2575	0.0000	2.000
Propyl	3807	6.3584	3.000	9.0091	0.0000	42.000